ROI sensitive analysis for real time gender classification

RODRIGUES, Marcos <http://orcid.org/0000-0002-6083-1303>, KORMANN, Mariza and TOMEK, Peter

Published version

RODRIGUES, Marcos, KORMANN, Mariza and TOMEK, Peter (2014). ROI sensitive analysis for real time gender classification. In: MASTORAKIS, Nikos, PSARRIS, Kleanthis, VACHTSEVANOS, George, DONDON, Philippe, MLADENOV, Valeri, BULUCEA, Aida, RUDA, Imre and MARTIN, Olga, (eds.) Advances in information sciences and applications : Proceedings of 18th International Conference on Computers (part of CSCC'14). Recent advances in computer engineering series, 1 (22). World Scientific and Engineering Academy and Society (WSEAS), 87-90.

Copyright and re-use policy

ROI Sensitive Analysis for Real Time Gender Classification

Marcos A Rodrigues, Mariza Kormann and Peter Tomek

***Abstract*—This paper addresses the issue of real time gender classification through texture analysis. The purpose is to perform sensitivity analysis over a number of ROI-Regions of Interest defined over face images. The determination of the smaller ROI yielding robust classification results will be used for fast computation of texture parameters allowing gender classification to operate in real-time. Results demonstrate that the ROI comprising the front and the region of the eyes is the most reliable achieving classification accuracy of 88% for both male and female subjects using raw data and non-optimised extraction and classification algorithms. This is a significant result that will drive future research on optimisation of texture extraction and linear discriminant algorithms.**

***Index Terms*—Face detection and tracking, texture analysis, gender classification**

I. Introduction

IMAGE analysis for gender classification has a number of useful applications such as collecting demographic statistics for marketing purposes, security surveillance, and the development of customised human-computer interfaces. The ADMOS project [1] is funded by the EC and aims to develop a real-time gender and age recognition to be used in private spaces of public use, such as shopping malls, fairs and outdoor events.

The purpose of this paper is to report on sensitivity analysis and performance evaluation concerning gender classification using texture analysis. The parameters under investigation include the size of the detected region in the image in relation to the location of the eyes, and various regions of the face. Specifically these include a larger image with aspect ratio *width*:*height* of $3:4$ which normally includes ears, hair and a portion of the neck, $1:1$ which is narrowed down to the region of the face from the front to the chin and normally includes portions of the ears, and $2:1$ which are defined as the top half of the face (front, eyes, portions of the nose and ears) and bottom half (including lips and portions of nose, chin and ears).

A set of experiments are designed to determine the optimal region of the face; ideally this would be as small as possible to allow the system to operate in real time. The techniques under investigation are the LBP–Linear Binary Pattern algorithm [2] in conjunction with eigenvector decomposition to determine class membership [3]. No LBP improvements are proposed here neither the optimisation of classification algorithms; the aim is to robustly assess the various ROI-Regions of Interest of an image. Once such regions are determined, then improvements to the techniques are investigated.

LBP is a non-parametric method used to summarise local structures of an image and have been extensively been exploited in face analysis for gender, age, and face recognition [4], [5], [6], [7], [8], [9]. Normally, LBP are employed in local and holistic approaches and a number of extensions have been demonstrated in the literature (e.g. [6]) in connection with linear discriminant analysis and SVM-support vector machines.

The approach in this paper is demonstrated by using a public database from which the various regions are automatically selected by face, eyes and

Marcos A Rodrigues and Mariza Korman are with the GMPR–Geometric Modelling and Pattern Recognition Research Group at Sheffield Hallam University, Sheffield, UK. Email {m.rodrigues, m.kormann}@shu.ac.uk

Peter Tomek is with ATEKNEA, Budapest, Hungary. Email peter.tomek@ateknea.com

This project has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no 315525.

lip detection algorithms. The method is described in Section II, experimental results are presented in Section III with a conclusion in Section IV.

II. Method

In order to perform real-time gender classification, a number of steps are necessary. First faces must be detected in the image and this is achieved through the Viola-Jones method [10], [11] available from OpenCV libraries. An unconstrained image may contain a number of faces and each region of interest ROI containing a face must be processed independently and the detected gender must be assigned to a corresponding data structure (to that region of interest). The data structure thus, will contain a gender attribute such that to avoid unnecessary multiple calls to the gender classification function if a particular tracked face has already a gender definition. This applies to the tracking of faces over multiple frames, but these aspects are not discussed further in this paper. Instead the focus is on sensitivity analysis over selected regions of the face.

Sensitivity analysis starts with testing aspect ratios of the ROI concerning the selected facial region. The Viola-Jones method yields a ROI with aspect ratio of $1:1$. The method, however, suffers from false positive detection: due to the simplicity of the Haar-like features used, some of the detected face objects are not faces at all. In order to verify whether or not it is a face, a number of constraints are imposed namely, a face must have two eyes and a lip. To verify the constraints each ROI is taken in turn. First eye detection is performed in the knowledge that right and left eyes are located in the first and second quadrants of the ROI, and lip detection in the knowledge that the lip is spread over the third and fourth quadrants. An example of a verified face is depicted in Figure 1. If the ROI satisfies these constraints then proceed to gender classification. Note that eye detection is a specific requirement of the ADMOS project but very useful in the context of sensitivity analysis presented here as the size of a larger ROI (of aspect ratio of 3:4) is determined from the eye locations.

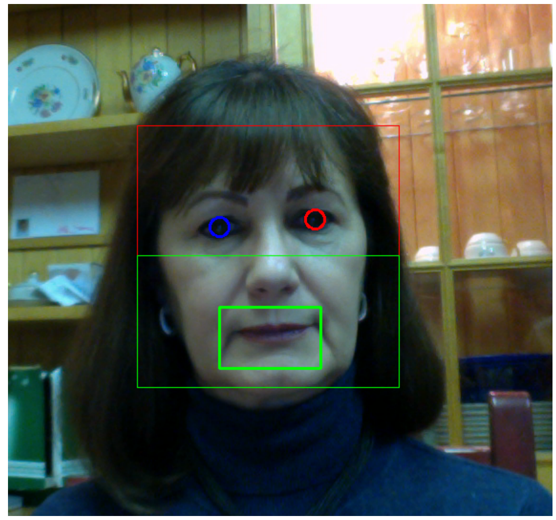


Fig. 1. Face, eyes, and lip detection in real time with annotated face ROI of aspect ratio $1:1$

Local binary patterns [2], [4] are grey-scale operators useful for texture classification defined over local neighbourhood pixels. It was originally defined using a 3×3 array of pixels. The value of the centre pixel is compared with its neighbours and the result (greater or smaller) expressed as a binary number and summed over all pixels considered. LBP can be expressed over P sampling points on a circle of radius R where the value of the centre pixel (x, y) is expressed as:

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} T(I_p - I_c) 2^p, \qquad (1)$$

where I_p and I_c refer to the pixel intensity in the grey level of the centre pixel and of P pixels on a circle of radius R, and T is a thresholding function with $T(.) = 1$ if $(I_p - I_c) \geq 0$ or $T(.) = 0$ otherwise. Normally images are defined in blocks from which individual LBPs are calculated and then concatenated into a single histogram. The analysis of such histograms can be used to differentiate texture patterns. The size of the block under analysis can vary and this obviously will be reflected in the LBP histogram.

In order to improve the robustness and discriminative power of the basic LBP operator as defined in equation (1) a number of variations to LBP have been proposed in the literature [5]. Note that the purpose of this paper is not to investigate possible extensions to LBP but to perform a sensitivity analysis to determine which region of the face is the most reliable for gender classification. Once this is achieved the most promising region will be investigated with a number of extensions to LBP and linear discriminant algorithms [3] including perceptron, relaxation rules, Fisher's criterion, least mean squared methods, and support vector machines as reported in the literature (e.g. [5], [6], [7]).

The approach to gender classification applied to facial regions can thus be stated as:

1) Define a set of stable measures $m_i(i = 1, 2, \ldots, n)$ on an image ROI and build a vector $M = (m_1, m_2, \ldots, m_n)^T$ that characterises the ROI;
2) Build a matrix Ω of vectors M where the index of M points to the identity of the ROI object: $\Omega = (M_1, M_2, \ldots, M_s)^T$ where s is the total number of images or vectors in the database;
3) Define a method to estimate the closest distance from a given ROI vector M to the most similar vectors in the database. The class of those vectors will point to the most likely class of M.

The set of stable measures is the histogram produced as a result of applying LBP to the data. It subsumes a large number of variables on the image ROI and the purpose is to identify which variables are more relevant to the problem at hand. Here the problem is approached by analysis of variance, or principal values. There are at least three ways a set of principal components can be derived. If m_i is the set of original variables, then they can be expressed as a linear combination ψ:

$$\psi_i = \sum_{j=1}^{p} a_{ij} m_j \tag{2}$$

The Hotelling approach described in [3] is used here in which the purpose is to find a linear separation for which the sum of the squares of perpendicular distances is a maximum, that is the choice is to maximise the variance of ψ. This is a common approach to the problem based on first determining the scatter matrix S by subtracting all values from their mean. Then the covariance matrix is estimated as $\Sigma = SS^T$. The principal components are determined by performing an eigenvector decomposition of the symmetric positive definite matrix Σ and then using the eigenvectors as coefficients in a linear combination of the original variables m.

In order to determine class membership, the method is to define a training set in which ground truth representative of male and female images are used and classified accordingly. The leave-one-out technique is used at testing stage. The principle is that the closest vector in the database real class is the proposed class for the testing vector. In order to improve accuracy, a voting method is proposed as the best of 5 closest matches – that is, for an unknown vector the proposed class is determined by majority counting.

III. Experimental results

A public database is used containing a large variation of faces concerning pose and illumination, details can be found in [12]. The database provides sets of large images which are not restricted to the facial area, and subjects are presented in a number of different poses other than frontal. Examples of images from the database are depicted in Figure 2.

A set of five experimental tests were carried out as follows.

1) ROI1: using a large ROI that includes hair, ears, and areas of the neck. This is selected from face detection and by adjusting the width and height of ROI with aspect ratio 3:4 based on the detected distance between eyes. For instance, if the inter-ocular distance is d, then the ROI width is defined with d to either side of the eyes, and $2d$ to the top and $2d$ to the bottom.
2) ROI2: using a ROI of 1:1 as detected by the face detection algorithms. This is narrowed down to the area of the face as compared

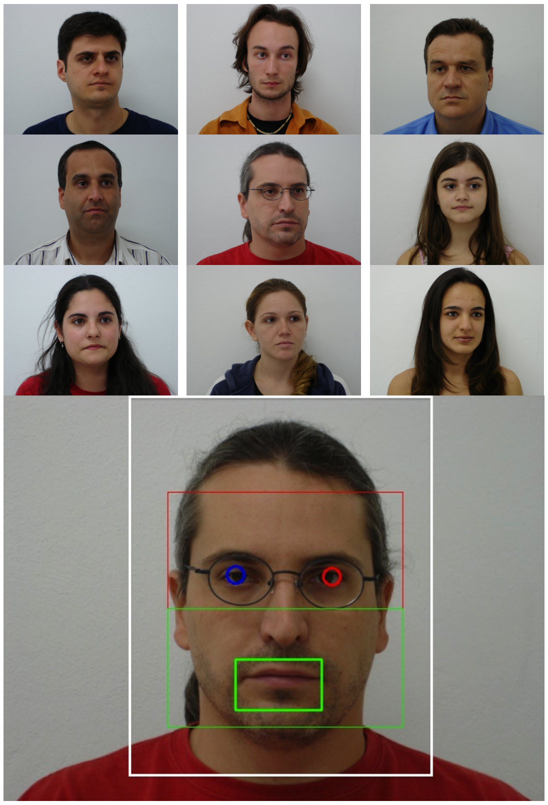


Fig. 2. Examples of data from the FEI database. The larger image shows the 5 automatically detected regions of interest.

to the previous larger ROI, and normally includes the areas from the front to the chin.

3) ROI3: using a ROI of 2:1 including only the top half of the detected face.
4) ROI4: using a ROI of 2:1 including only the bottom half of the detected face.
5) ROI5: using a ROI of 2:1 corresponding to the detected area around the lips.

A set of 100 images were used, 50 male and 50 female. For each ROI, LBP was performed to extract the relevant features represented by the histogram. An example is shown in Figure 3. The histograms for each ROI are used for training and testing purposes. Each ROI was tested independently as the purpose is to determine which one yields more accurate results; specifically the smaller effective ROI is sought as the purpose is to be applied in real time gender classification.

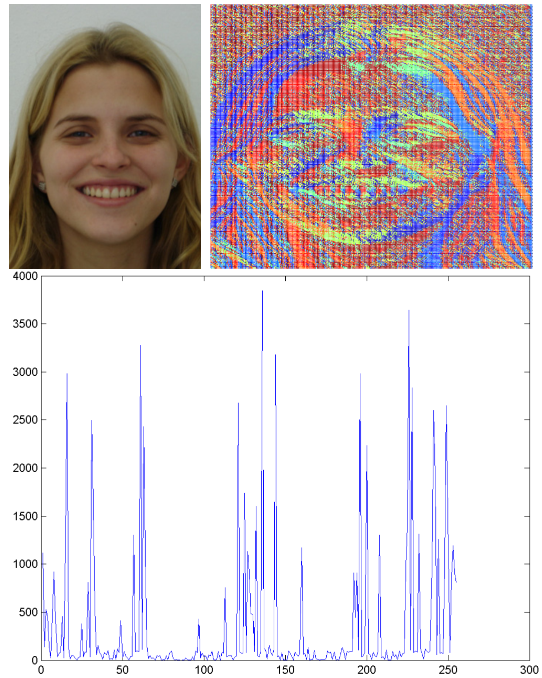
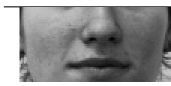


Fig. 3. Example of LBP and histogram for a selected ROI.

The implementation was carried out in Matlab following the method described in Section II. Results were automatically saved to a spreadsheet for further analysis. Performance in terms of accuracy were noted and results are summarised in Table I. Results for ROI1 yielded the highest accuracy for male classification at 92%, and female classification was substantially lower at 79%. This is probably because adding portions of background together with large amounts of hair increases the overall variance which can be confused with skin variations of male subjects due to facial hair. In any case, the inclusion of background is not desirable as it can introduce errors that cannot be controlled.

Results for ROI2 were slightly improved for female subjects at 83% but worst for male ones at 83%. Still, results are more consistent and thus, preferable to ROI1. ROI3 (consisting of the upper half of ROI2) yielded the best results, at 88% accuracy for both male and female subjects. This shows that there is enough variation in this region

TABLE I
COMPARATIVE ANALYSIS OF IMAGE REGIONS

Image ROI	Gender	Classification results
 ROI1	Male Female	92% 79%
 ROI2	Male Female	83% 83%
 ROI3	Male Female	**88%** **88%**
 ROI4	Male Female	88% 71%
 ROI5	Male Female	<40% <40%

to detect both classes with good accuracy. ROI4 (consisting of the bottom half of ROI2) retained its previous accuracy for male subjects at 88% but for female the performance decreased substantially to 71%. This shows that the region does not contain enough variation in texture to guarantee robust classification, which is somewhat surprising. Before testing this ROI, it was expected that, due to facial hair and skin texture variations on either shaved and unshaved faces, male and female subjects would be detected with the highest degree of accuracy. Finally ROI5 has proved to be problematic as automatic lip detection failed in about half of the data (multiple lip detection was deemed a failure). Thus the lack of robust lip detection rendered this ROI unstable and the statistics reported here at less than 40% are just an indication of performance.

Since this research is only interested in the minimum ROI with highest performance, results demonstrate that ROI3 is the most promising one. The high variance caused by skin texture, eyebrows, eyes, and portions of hair mean that the region can be further exploited for more robust classification. With this information, it is now possible to focus on improvements to LBP and to the classification algorithms. The obvious avenues to explore include the use the Fisher's criterion and SVM-support vector machines. In particular, SVM holds the highest promise as it is a technique designed to maximising the margin between canonical hyperplanes separating the two classes. With a higher margin, the probability of misclassification naturally decreases and it should be possible to achieve higher accuracy for both male and female subjects.

IV. CONCLUSION

This paper has presented a sensitivity analysis of gender classification based on LBP and eigenvector decomposition over 5 regions of interest. The adopted methodology was directed towards testing the LBP algorithm over the selected regions to determine the smaller region of interest yielding the best classification results. Images from a public database were used on which automatic detection of the face region, eyes, and lips were performed resulting in the various regions of interest being defined. Classification results show that the region containing parts of the nose, eyes and front are the most reliable, with an overall accuracy of 88% for both male and female subjects.

This is a significant result on its own right but, more importantly, it provides a focus on which to apply optimisation techniques to bring overall accuracy to a desirable level of 95%. Other techniques reported in the literature such as in [6] perform specific selection of histogram bins and use SVM for feature classification and are shown to perform slightly better than using the raw LBP histogram as reported here. Those results and the results obtained in this paper clearly indicate that, by using the selected region and developing similar techniques it will be possible to substantially increase classification accuracy. Furthermore, by allowing the use of a substantially smaller region as compared to the ones reported in the literature, computation will be much faster and appropriate for real-time applications. Research is under way on fine-tuning LBP for the selected region through the use of SVM and related methods and results will be reported in the near future.

REFERENCES

[1] ADMOS (2013) *Advertising Monitoring System Development for Outdoor Media Analytics*, EC Grant Agreement 31552. [Online] Available at http://admos.eu

[2] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 28(12):20372041, 2006.

[3] A. Webb and K. Copsey (2011) *Statistical Pattern Recognition*, 3rd edition, Wiley, 666pp.

[4] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen. (2011) Computer Vision Using Local Binary Patterns. *Springer*.

[5] Ylioinas, J., Hadid, A., Pietikäinen, M. (2011) Combining contrast and local binary patterns for gender classification. *SCIA 17th Scandinavian Conference on Image Analysis.*

[6] C. Shan (2012) Learning local binary patterns for gender classification on real-world face images, Patter Recognition Letters 33 (2012) 431–437.

[7] Y. Guo, G. Zhao, M. Pietikäinen, and Z. Xu. Descriptor learning based on fisher separation criterion for texture classification. *In Proc. ACCV10*, 185–198, 2010.

[8] Lian, H., Lu, B. (2007) Multi-view gender classification using multi-resolution local binary patterns and support vector machines. *Int J Neural Systems* 17 (6), 479–487

[9] Sun, N., Zheng, W., Sun, C., Zou, C., Zhao, L. (2006) Gender classification based on boosting local binary pattern. *In: Int Symp on Neural Networks.*

[10] Viola, P., Jones, M. (2001) Rapid object detection using a boosted cascade of simple features. *In: IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR), pp. 511518.

[11] Viola, P., Jones, M. (2004) Robust real-time face detection. *Internat. J. Comput. Vision* 57 (2), 137154.

[12] (2014) FEI Face Database. [Online] Available http://fei.edu.br/ cet/facedatabase.html
(2011)